



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

OPTIMIZING UTILIZATION OF DETECTORS

by

Crystal R. Warrene

March 2016

Thesis Advisor:
Second Reader:

Frank E. Kragh
Daniel Bursch

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE March 2016	3. REPORT TYPE AND DATES COVERED Master's thesis		
4. TITLE AND SUBTITLE OPTIMIZING UTILIZATION OF DETECTORS			5. FUNDING NUMBERS	
6. AUTHOR(S) Crystal R. Warren				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) <p>This work seeks to increase the expected intelligence value collected by optimizing the time on multiple tasks. The purpose of this thesis is to provide a quantifiable process to determine how much time should be allocated to each task sharing the same asset.</p> <p>This optimized expected time allocation is calculated by numerical analysis and Monte Carlo simulation. Numerical analysis determines the expectation by involving an integral and a joint probability density function for a range of rates. In this case, rates are the historical hailing by taxi passengers. Monte Carlo simulation determines the optimum time allocation of the asset by repeatedly running experiments to approximate the expectation of the random variables. This was deemed necessary to account for real-world uncertainties as applied to a taxi scenario. The taxi variables consist of hail rates of the passengers, the fare amount for the task, and how much time to pursue said fare. Accounting for the uncertainty in the hail rates was exhibited by using ranges and not given values. The relationship the rates of hails for the taxi from two passengers and the fare values gathered is important to utilizing the taxi to maximize the total fare collected.</p>				
14. SUBJECT TERMS optimizing time, Monte Carlo simulation, Numerical analysis, space asset, asset usage			15. NUMBER OF PAGES 79	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

OPTIMIZING UTILIZATION OF DETECTORS

Crystal R. Warrene
Lieutenant, United States Navy
B.S., United States Naval Academy, 2009

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN SPACE SYSTEMS OPERATIONS

from the

**NAVAL POSTGRADUATE SCHOOL
March 2016**

Approved by: Frank E. Kragh, Ph.D.
Thesis Advisor

Daniel Bursch,
Second Reader

Rudy Panholzer, Ph.D.
Chair, Space Systems Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This work seeks to increase the expected intelligence value collected by optimizing the time on multiple tasks. The purpose of this thesis is to provide a quantifiable process to determine how much time should be allocated to each task sharing the same asset.

This optimized expected time allocation is calculated by numerical analysis and Monte Carlo simulation. Numerical analysis determines the expectation by involving an integral and a joint probability density function for a range of rates. In this case, rates are the historical hailing by taxi passengers. Monte Carlo simulation determines the optimum time allocation of the asset by repeatedly running experiments to approximate the expectation of the random variables. This was deemed necessary to account for real-world uncertainties as applied to a taxi scenario. The taxi variables consist of hail rates of the passengers, the fare amount for the task, and how much time to pursue said fare. Accounting for the uncertainty in the hail rates was exhibited by using ranges and not given values. The relationship the rates of hails for the taxi from two passengers and the fare values gathered is important to utilizing the taxi to maximize the total fare collected.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	MOTIVATION	1
B.	UNIQUE FOCUS OF THIS THESIS	2
C.	LITERATURE REVIEW	3
1.	Statistical Concepts.....	3
2.	Application of Statistical Concepts.....	4
D.	OVERVIEW	7
II.	FUNDAMENTAL STATISTICAL CONCEPTS	9
A.	DISTRIBUTION FUNCTIONS	9
B.	EXPECTATION	11
C.	DISTRIBUTIONS.....	12
D.	OVERVIEW.....	13
III.	METHODOLOGY	15
A.	NUMERICAL ANALYSIS	15
B.	MONTE CARLO SIMULATION.....	18
C.	OVERVIEW	20
IV.	RESULTS	21
A.	COMPARISON OF RESULTS.....	21
1.	Group One	22
a.	<i>Scenario One</i>	<i>24</i>
b.	<i>Scenario Two and Three.....</i>	<i>25</i>
c.	<i>Scenarios Seven and Eight</i>	<i>28</i>
d.	<i>Scenarios Four through Six</i>	<i>29</i>
2.	Group Two.....	31
3.	Group Three and Group Four.....	32
a.	<i>Scenario 18.....</i>	<i>34</i>
4.	Group Five.....	35
a.	<i>Table 5 Compared to Table 3.....</i>	<i>36</i>
b.	<i>Scenario 36.....</i>	<i>37</i>
5.	Group Six.....	38
a.	<i>Effects of a Close Range of Hail Rates</i>	<i>39</i>
b.	<i>Scenario 45.....</i>	<i>39</i>
6.	Group Seven	40
a.	<i>Scenario 51 and 52.....</i>	<i>41</i>

<i>b.</i>	<i>Table 7 versus Table 6</i>	42
<i>c.</i>	<i>Scenario 49</i>	42
B.	OVERVIEW	43
V.	SUMMARY OF FINDINGS AND AVENUES FOR FURTHER RESEARCH	45
A.	CONCLUSION	45
B.	FURTHER RESEARCH	46
	APPENDIX A. CLASSIFIED CHAPTER	47
	APPENDIX B. EQUATION (1.1) AND (1.2) DERIVATION	49
	APPENDIX C. MATLAB CODE FOR NUMERICAL ANALYSIS	53
	APPENDIX D. MATLAB CODE FOR MONTE CARLO SIMULATION	55
	APPENDIX E. PROBABILITY THAT HAIL RATE OF PASSENGER ONE IS LARGER THAN THE HAIL RATE OF PASSENGER TWO	59
	LIST OF REFERENCES	61
	INITIAL DISTRIBUTION LIST	63

LIST OF FIGURES

Figure 1.	Visual representation of a limited resource, the taxi, with multiple tasking.	3
Figure 2.	Cumulative distribution function for a uniform random variable [10].	10
Figure 3.	Scenario One histogram.	25
Figure 4.	Scenario Two histogram.	27
Figure 5.	Scenario Seven histogram.	29
Figure 6.	Scenario Four histogram.	31
Figure 7.	Scenario 10 histogram.	32
Figure 8.	Scenario 18 histogram.	34
Figure 9.	Scenario 45 histogram.	40
Figure 10.	Scenario 49 histogram.	43

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	Results for numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various uniform distributions for the hail rate for Passenger One, R_1	24
Table 2.	Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various uniform distributions of hail rate for Passenger Two, R_2	32
Table 3.	Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various conditional fare values for Passenger One, I_1	33
Table 4.	Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various conditional fare values for Passenger Two, I_2	35
Table 5.	Resultant expected optimum time for Passenger One, $E(T_1)$, for various conditional fare values for Passenger One, I_1 , and different hail rate for Passenger One, R_1 , and Passenger Two, R_2	36
Table 6.	Resultant expected optimum time pursuing Passenger One, $E(T_1)$ for various conditional fare values for Passenger One, I_1 and $\Pr(R_1 > R_2) \ll 0.5$	38
Table 7.	Resultant expected optimum time pursuing Passenger One, $E(T_1)$ for various conditional fare values for Passenger One, I_1 , combined with a much smaller hail rate for Passenger One, R_1	41

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would like to thank Dr. Frank Kragh for his endless patience and willingness to ensure that I understood the research process and the mathematical concepts. His accommodating, easygoing nature was refreshing. I appreciate his guidance regarding the subject of this thesis, the insights into Monterey life, and the information about how to publish later on in life.

I also would like to express my gratitude to my friend Cole—an Air Force flight engineer who does not judge me for wearing a non-Air Force uniform—for the MATLAB tricks and skills he shared.

Finally, to my family members who put up with the long nights, weekends away, and frustrated contemplations. I could not do this job without you. Please know that you are my motivation to give my absolute best to the country we enjoy and live in every day.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Americans entrust the government to manage their hard-earned tax dollars effectively. In this difficult economic climate, matters of financial efficiency deserve even greater attention. For example, UPS is deploying an on-road integrated optimization navigation (ORION) system to “select the most efficient route while meeting customer requirements” [1]. The ORION system is predicted to reduce fuel cost, vehicle maintenance cost, and CO₂ emissions. Optimization tools, such as ORION, allow decision makers to use existing systems more efficiently, resulting in cost savings by requiring fewer assets to meet customer demands. Furthermore, these tools often allow the collection of data that otherwise would have been neglected. This efficiency in time allocation is important to decision makers, mission planners, and operators because it may increase the overall value of the intelligence that is gathered. Optimizing the use of our national assets ensures due diligence on behalf of the taxpayers, more intelligence value for the money spent by the U.S. government, and the possibility of collecting more data to make informed decisions.

A. MOTIVATION

If multiple tasks require a common government resource, what is the optimal way to allocate the time spent on that resource? Decision makers need a scientific and quantifiable process to determine the best way to allocate resources for an asset’s tasks. In this thesis, we examine a method for using limited resources to their maximum intelligence-gathering potential. This quantitative process is a tool that allows decision makers to be aware of the effect their decisions may have on the amount of intelligence gathered by the resource. The approach seeks to increase the expected intelligence value collected by an asset—not by tapping its existing potential nor making changes to its structure,—but instead by improving the tasking for that resource. More effective time allocation and planning may assist in realizing maximum expected potential and increasing the quality of service. This research may result in further consideration of the relationship between time allocation and the overall capacity of the resource. The purpose

of this research is to provide information on the use of limited national assets to increase the overall intelligence value gathered.

B. UNIQUE FOCUS OF THIS THESIS

With random variables, this thesis will use an expected estimate of intelligence gathered to optimize the efficiency of very expensive, limited national assets. It is important to find the optimal time for a resource to accomplish each of its tasks. The more efficient use of the resource; the more valuable the intelligence gathered.

Numerical analysis and computer simulation of select random variables provide a method of determining how to use limited national assets to maximize time optimization. This research focuses on one type of limited asset, which is discussed in classified Appendix A. However, for the purpose of this thesis, I used a hypothetical unclassified example of a taxi with two tasks.

This approach assigns cab fares to one taxi, which must pick up two separate passengers to complete its mission. The taxi is considered the limited resource, and the taxi scenario is affected by how much fare is charged (i.e., the fare value, and how much time is allotted to picking up the passengers). The research focus is on optimizing the division of a fixed total time into the time to find Passenger One and the time to find Passenger Two, with respect to the average rates of finding Passenger One and Passenger Two. Each rate is defined as how fast or slow the taxi is estimated to be hailed by a passenger, measured in hails per hour, or simply per hour. Each rate is independent of the other. For example, based on historical data, Passenger One may be predicted to take a long time to hail a cab. Each passenger represents a task of the taxi. The question becomes, “What time should be allocated to pick-up Passenger One and Passenger Two to maximize the total expected fare?” The goal of this research is to optimize the time spent on collecting each of the two passengers jointly, thereby maximizing the expected total fare collected, measured in dollars. In the scenario of a taxi picking up Passenger One, optimization of time allocation affects the ability of the taxi to accomplish Task Two (See Figure 1). This thesis examines time optimization given the pair of joint tasks, conditional fare values, and the average rates of each task.



Figure 1. Visual representation of a limited resource, the taxi, with multiple tasking.

Adapted from [2]: “Map,” n.d. Google [Online]. Available at: <http://images.google.com/>. [Accessed: 29-Jan-2016]. and [3] taxi image “Taxi,” *Dreamstime*. [Online]. Available at: <http://www.dreamstime.com/royalty-free-stock-images-cartoon-yellow-taxi-car-image21652969>

C. LITERATURE REVIEW

Other scholars have written about the same concepts that are discussed throughout this thesis, providing a foundation for the methods used here. This section is divided into information found in introductory statistics textbooks and previous theses that apply these statistical concepts. The introductory section examines random variables and key probability approaches. The previous theses section discusses algorithms to improve the usage of assets.

1. Statistical Concepts

Random variables measure chance events associated with a sample space and can be defined as a numerically valued function over a sample space [4]. Continuous random variables are critical to this investigation. In the taxi analogy, we may model the rate to find each passenger as a random variable with perhaps infinitely many possible rates to collect Passenger One and Passenger Two. Modeling these rates as random variables makes sense if we do not know these rates precisely. These random variables and how

they impact the optimum taxi allocation can be analyzed using well-established analytical statistical methods. These methods allow the calculation of theoretical probabilities and averages using calculus and algebra. The act of collecting, analyzing, interpreting, and drawing conclusions about the numerical data is called statistics [5].

Another approach used in this research is correlating probabilities to real-world events with relative frequency [5]. Relative frequency involves a large number of trials to reveal the probability of an event and the relative frequency of its occurrence [5]. Testing an experiment a large number of times under identical environmental conditions creates an empirical probability [6]. There may be a difference between empirical probability and theoretical probability. However, if the experiment is run a very large number of times, the empirical probability, and theoretical probability will be approximately the same. According to the Law of Large Numbers [5], the difference between the empirical probability and the theoretical probability can be made arbitrarily small by increasing the number of trials. In this research, computers were used to run a large numbers of trials. Empirical probability is limited by the fact that the results are hypothetical and that careful consideration is needed when pinpointing the number of trials [6]. Details of these simulation concepts are discussed in Chapter II.

2. Application of Statistical Concepts

Previous work provides the background needed to understand the stakes and the application of the optimization of national assets. For example, in *Dynamic Scan Schedules*, Dutertre [7] maximized the quality of service metric for scheduling equipment usage on airplanes, specifically emitter finders. Dutertre applied an algorithm to improve asset usage and improving asset usage in a dynamic environment is the focus of this thesis. Moreover, this thesis builds on work by Chris M. Duke [8] and Kenneth St. Germain [9], with a particular emphasis on St. Germain's time optimization equation for the time allocation between two tasks time-sharing the same asset.

An article by Dutertre [7] shows how an algorithm with guaranteed detection probabilities was applied to enhance the scan schedules for airplanes. He did this to find the optimal schedule to optimize the quality-of-service metric. Dutertre declares his

“improvements were demonstrated via simulation, but the basic techniques can be extended and generalized for even better performance” of the scanners. His algorithm makes the schedule for the signal emission scanning in real time and improves the detection of the signal of interest performance as search preferences adjust

Duke examined how our nation currently tasks limited resources in his masters’ thesis *Optimizing the Signals Intelligence Tasking Process* [8]. Duke visited organizations and studied how those organizations tasked their limited resources. Building on Duke’s thesis, St. Germain presented a quantifiable optimization methodology and a metric that can be used to enhance the efficiency of intelligence collecting for national resources [9]. His effort to calculate the average intelligence was determined by the probability of collection using assumed deterministic constants for conditional intelligence value and rates of collections using Equations (1.1) and (1.2). The key result of St. Germain’s thesis is

$$T_1 = \begin{cases} 0 & \text{if } T < 0 \\ T & \text{if } 0 \leq T \leq T_{Total} \\ T_{Total} & \text{if } T > T_{Total} \end{cases} \quad (1.1)$$

where

$$T = \frac{\left[\ln \left(\frac{I_1 R_1}{I_2 R_2} \right) + R_2 T_{total} \right]}{(R_1 + R_2)} \quad (1.2)$$

which is derived in Appendix B and where I_1 is the conditional intelligence value of Task One or the conditional fare charged for Passenger One. I_2 is the conditional fare charged for Passenger Two. R_1 is the average rate of collection for Task One or the rate at which the taxi is hailed by Passenger One. R_2 is the rate at which the taxi is hailed by Passenger Two. The variables T_{total} is the total time allotted for the resource to attempt both tasks, and T_1 is the expected optimum time spent pursuing Passenger One.

Using Equations (1.1) and (1.2), the resource management team of mission planners and operators could make an informed decision with regard to resource allocation and expected intelligence value gained or lost, thereby increasing, on average,

the intelligence gathered by assets. In St. Germain's thesis [9], the optimum time was the only random variable assigned to a resource to accomplish multiple taskings from the perspective of a mission planner at a specific mission control station.

This research will re-analyze the optimum time problem with use of random variables for the average rates of collection, R_1 and R_2 , and the conditional intelligence values, I_1 and I_2 , to accomplish multiple tasks. In previous work by Duke [8] and St. Germain [9], they used constant values for the rates of collection and conditional intelligence. One can apply historical data to determine an approximate average of collection rates for various signals of interest in specified geographical areas at specified times of the day and year. The same historical data can create approximate histograms, or effectively probability density functions, for the collection rates. Since we have a finite amount of historical data, we do not know the precise value for average rate of collection, so it would be more accurate to treat the average rates of collection, R_1 and R_2 , as random variables, where the range of the random variables includes all the values of average collection deemed likely, according to the historical data. The optimum time allotted will be calculated both via numerical analysis and separately by Monte Carlo simulation. If these two methods yield the same results, we can have high confidence that the results are correct.

This work accounts for the uncertainty and variabilities in average collection rates by treating St. Germain's result (Equations (1.1) and (1.2)) as a conditional expected value of the optimum time allocated to Task One conditioned on the average rates of collection. Then this work calculates the expected value of the optimum time allotted by removing the conditions on average rates of collection, i.e., by averaging over the joint distribution for the two average collection rates. Thus, this thesis generates estimates for optimized allotted time that is not sensitive to assumed values for the average rates of collection.

D. OVERVIEW

The remainder of this thesis is organized as follows: Chapter II examines mathematical concepts that must be understood to comprehend methods in later chapters and the value of the results. Chapter III illustrates how two methods—numerical analysis and Monte Carlo simulation use Equations (1.1) and (1.2) to come to the same expected optimum time and why two approaches are desired for this research. Chapter IV reviews the findings and behaviors of the numerical analysis and Monte Carlo simulation, and Chapter V discusses conclusions and suggests further work.

THIS PAGE INTENTIONALLY LEFT BLANK

II. FUNDAMENTAL STATISTICAL CONCEPTS

This chapter explains key mathematical computations used in this study. The chapter begins with basic concepts in statistics used to manipulate Equations (1.1) and (1.2). These manipulations will be applied in Chapter III. Fundamental concepts of statistics and probability that are necessary for understanding this thesis fall into three categories: distribution functions, expectation, and random variable types. These distributions include the cumulative distribution function, probability distribution function, expectation, Poisson distribution, uniform distribution, and exponential distribution. The distribution functions allow Equations (1.1) and (1.2) to be applied differently from previous works. The distributions affect the possible random variable sample sets to be used as conditional rates and as conditional intelligence values.

A. DISTRIBUTION FUNCTIONS

This section explains how Equations (1.1) and (1.2) can be manipulated and used in various ways when compared to the previous works of Duke [8] and St. Germain [9]. First, the cumulative distribution function (CDF) involves probability and random variables in one simple equation. The derivative of the CDF produces a probability density function (PDF).

The CDF is defined mathematically as

$$F_X(x) \equiv \Pr(X \leq x) \quad (2.1)$$

where X is a random variable and x is any possible value of X [6]. The CDF is a method of describing how the possible values of the random variables are distributed and can be applied to random variables of any type [6]. A CDF for a discrete random variable consists of a number of discontinuous steps, whereas a CDF for a continuous random variable is like the graphs shown in Figure 2, which are monotonically increasing [6]. Monotonically increasing simply means that the function never has a negative derivative [6]. The CDF of a uniform random variable is linear, as shown in Figure 2 [6].

The maximum value of the CDF is one, $\lim_{x \rightarrow +\infty} F_X(x) = 1$; this knowledge can be used to eliminate variables in Equation (1.2) [6].

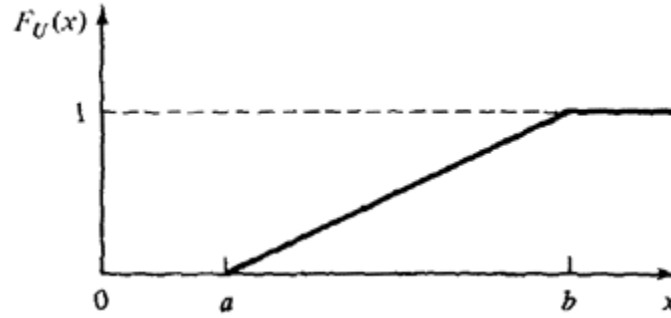


Figure 2. Cumulative distribution function for a uniform random variable [10].

Source: [10] R. C. Larson and A. Odoni, "2.10 Probability Density Functions," [Online].
Available at: http://web.mit.edu/urban_or_book/www/book/chapter2/2.10.html.

The CDF is always monotonically increasing, which means $F_X(x_1) \geq F_X(x_2)$ if and only if $x_1 \geq x_2$ [6]. A simple example to illustrate this point is the probability that a freezer's temperature is less than or equal to negative four degrees, i.e., $F_Y(-4^\circ) = \Pr(Y \leq -4^\circ) = 1/3$, and the probability that the freezer's temperature is less than or equal to negative five degrees, i.e., $F_Y(-5^\circ) = \Pr(Y \leq -5^\circ) = 1/2$. Notice that this is impossible because the probability that the temperature is less than -4° cannot be less than the probability that the temperature is less than -5° , i.e., $\Pr\{Y \leq -4^\circ\} = \Pr\{Y \leq -5^\circ\} + \Pr\{-5^\circ < Y \leq -4^\circ\} \geq \Pr\{Y \leq -5^\circ\}$. This example would be graphed with a negative slope, violating the rule of being monotonically increasing and demonstrating that for a CDF, a negative slope is impossible [6]. Joint CDFs must also be discussed due to the number of variables in Equations (1.1) and (1.2). Instead of $F_X(x)$, a joint CDF is written as $F_{X,Y}(x, y) \equiv \Pr(X \leq x, \text{ AND } Y \leq y)$ [6]. It is also important to understand the relationship between the CDF and PDF.

The PDF is the derivative of the CDF; mathematically defined as [6]

$$f_x(x) = \frac{d}{dx} F_x(x) \text{ or } f_x(x) = F'_x(x). \quad (2.2)$$

This relationship can be inverted so that the CDF can be calculated by using the area under the PDF, in other words,

$$F_x(x) = \int_{-\infty}^x f_x(\alpha) d\alpha. \quad (2.3)$$

For the purposes of this thesis, it is further necessary to understand the conditional PDF with an emphasis on two random variables, R_1 and R_2 . The word conditional implies additional limitations or restrictions on the set of probabilities [6]. For example, how often would someone see a hummingbird? A conditional factor could be restricting the time duration to nighttime. Another important aspect of a PDF is the joint probability density function [4], that is

$$f_{A,B}(a,b) \equiv \frac{\partial}{\partial a} \frac{\partial}{\partial b} F_{A,B}(a,b) \quad (2.4)$$

which is used later in this chapter to find the expectation of joint variables. To review, the joint CDF of A and B is [5]

$$F_{A,B}(a,b) \equiv \Pr\{A \leq a \text{ AND } B \leq b\}, \quad (2.5)$$

and the joint PDF of A and B is Equation (2.4).

B. EXPECTATION

This section explains the expectation concept and how it will be used in Chapter III. The key point to remember is that the expectation is also the average. We can calculate the expected value using numerical analysis. Equations (1.1) and (1.2) will result in T_1 values that are averaged to calculate the expected optimum time. Another approach to determining the expectation is to use Monte Carlo simulation. Monte Carlo simulation is a theoretical approach used to determine the approximate expectation or average by repeating the experiment many, many times in the exact same environment [11].

The expectation is the mean or average, in this case, of the optimum time [4]. The mathematical process for calculating an average is different for discrete random variables and continuous random variables [6]. The mean of a continuous random variable, X will be examined. The expectation for a continuous random variable is the integral of the weighted averages times the values of the random variable. The PDF is essential because it is the weight applied to get the weighted average. The expectation of random variable

Y is mathematically defined as $E(Y) \equiv \int_{-\infty}^{\infty} y \underbrace{(f_Y(y)dy)}_{\substack{\text{the weights} \\ \text{the values} \\ \text{being} \\ \text{averaged}}} [6]$. The average must be

integrated with respect to the random variables. For a function of two joint continuous random variables, the average of the function $h(A, B)$ is

$E(h(A, B)) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(a, b) f_{AB}(a, b) da db [6]$, where f_{AB} is the joint probability density

function of A and B . If the two random variables are independent, this simplifies to

$$E(h(A, B)) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(a, b) f_A(a) da f_B(b) db \quad (2.6)$$

where f_A and f_B are the probability density functions for A and B , respectively [6].

Equation (2.6) will be seen later in Chapter III.

A further detail regarding dependent and independent variables must be emphasized. For two random variables to be called independent random variables, then the statistics of each random variable are the same whether or not the value of the other random variable is known. In this research, R_1 and R_2 are independent random variables and therefore have no effect on each other.

C. DISTRIBUTIONS

This section examines different distributions including Poisson and uniform. St. Germain's previous work explored Poisson's distribution that was used to derive Equations (1.1) and (1.2) [9] as seen in Appendix B. However, this study examined uniform distributions, as seen in Chapter III.

Previous work analyzed the probability of a number of discrete random events occurring in a fixed interval of time, modeled as a Poisson random variable [9]. The Poisson distribution applies when the time of each event is random with a constant average occurrence rate [9]. For example, the number of fares a taxi cab driver can collect in a specific number of hours could be modeled as a Poisson random variable. The probability mass function of a Poisson process is given by

$$p(x) \equiv \Pr(X = x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{if } x \in 0, 1, 2, 3, 4, \dots \\ 0 & \text{otherwise} \end{cases} \quad [9] \text{ where } X \text{ is the random variable, } x \text{ is}$$

the possible values of X , e is the base of the natural logarithm ($e = 2.71828 \dots$), and λ is the average number of events over the given time interval, T . Additionally, for $\lambda = RT$, R is the average rate of the taxi being hailed by passengers successfully per unit of time. The variable of rates is NOT based on the speed of the taxi. The probability mass

distribution function must sum to one, i.e., $\sum_{x=0}^{+\infty} \Pr(X = x) = \sum_{x=0}^{+\infty} \frac{e^{-\lambda} \lambda^x}{x!} = 1$ [9].

A uniform random variable is one that is equally likely to take on any value between two numbers. The ability to replicate real-world events by preventing bias from being introduced into the calculations gives creditability to this research. An example of a continuous uniform distribution applied to the previously mentioned taxi scenario is when the rate of taxi hailing is assumed to be unknown, but equally likely to be any value between one fare per hour and two fares per hour.

D. OVERVIEW

This chapter explained key concepts to include the definition of a CDF and how it can be transformed into a PDF. The CDF and the PDF must be applied to joint events, and a slight modification to the original forms of the CDF and PDF occurs when considering multiple joint events. This chapter also discusses the concept of and formula for the expectation so that its use in later chapters can be readily understood. Moreover, various random variable distributions are included because they are used in later chapters in this work.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

Based on introductory statistics concepts reviewed in Chapter II, this chapter discusses the manipulations of Equations (1.1) and (1.2) to find the optimum time allocation. The numerical analysis method uses the double integral of Equation (2.6) and a uniform distribution for random variables to find a value for the optimum time allocation of the asset. The Monte Carlo simulation method uses the relative frequency concept to determine the optimum time allocation of the asset by repeatedly running random experiments to approximate the expectation of the random variable T_1 in order to verify that the numerical analysis results are accurate. In this case, the Monte Carlo simulation results confirm the numerical analysis results. When numerical analysis and Monte Carlo simulation are applied to Equations (1.1) and (1.2), the findings are statistically identical and thus will be referred to as $E(T_1)$ for the rest of this thesis. Results for $E(T_1)$ are discussed in Chapter IV.

The two methods are required to replicate real-world uncertainty in Equations (1.1) and (1.2). This research is attempting to account for the unanticipated circumstances of real-life operations when finding the expected optimum T_1 . This objective is achieved by removing conditions via numerical analysis, by applying relative frequency in Monte Carlo simulation, and by comparing the two methods' results. Uncertainty must be accounted for with probability distributions and computational simulations to improve the accuracy of these results [12]. The two methods provide verification because if the two diverse methods yield the same answer, then both methods are likely correct; the answer, the expected optimum time to collect Passenger One, therefore is also correct.

A. NUMERICAL ANALYSIS

This section will illustrate how numerical analysis uses Equations (1.1) and (1.2) to find the expected optimum time allocation for the asset. Due to our assumed uncertainty in the rates, it is necessary to remove the conditions on the rates to use Equations (1.1) and (1.2) to find a value for the average optimum time.

This method builds on previous work [9], where the variables R_1 , R_2 , I_1 , I_2 , and T_{total} were considered deterministic constants that were used to determine an unconditional optimum T_1 . However, this study considers R_1 and R_2 as random variables, to account for real-world uncertainties. The introduction of random variables transforms Equations (1.1) and (1.2) into formulas that can calculate the conditional optimum time to be invested in Task One. We will use Equations (1.1) and (1.2), restated here, as the conditional expectation where the condition is that R_1 and R_2 are known

$$T_1(T) = \begin{cases} 0 & \text{if } T < 0 \\ T & \text{if } 0 \leq T \leq T_{Total} \\ T_{Total} & \text{if } T > T_{Total} \end{cases} \quad (3.1)$$

where

$$T(R_1, R_2) = \frac{\ln\left(\frac{I_1 R_1}{I_2 R_2}\right) + R_2 T_{total}}{R_1 + R_2} \quad (3.2)$$

where I_1 is the conditional fare charged for Passenger One and I_2 is the conditional fare charged for Passenger Two. Variable R_1 is the rate at which the taxi is hailed by Passenger One and R_2 is the rate at which the taxi is hailed by Passenger Two. T_{total} is the total time allotted for the resource to attempt both tasks, and T_1 is the expected optimum time spent pursuing Passenger One, as introduced in Chapter I.C.2. Note that the condition on each conditional fare charged is that the corresponding passenger was found and served. Thus, the conditional fare charged is fixed and does not account for the possibility that the passenger is not served.

We can calculate the expectation of T_1 by removing the conditions of the presumed values of R_1 and R_2 . This is the same as calculating the weighted average over all the possible values of random variables R_1 and R_2 where the weights are the joint PDF of R_1 and R_2 . Applying the average of the optimum time from Equation (3.1) and (3.2) with random variables, R_1 and R_2 , creates an average value for T_1 . Recall an expectation of a function of two random variables is calculated via Equation (2.6). Removing the

conditions on T_1 is the same as calculating an expectation of a function of two random variables. Applying Equation (2.6) from Chapter II to finding the expectation of T_1 yields

$$E(T_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_1(T(r_1, r_2)) f_{R_1 R_2}(r_1, r_2) dr_1 dr_2 \quad (3.3).$$

where $f_{R_1 R_2}$ is the joint PDF of R_1 and R_2 , the function T_1 is given in Equation (3.1), and the function $T(r_1, r_2)$ is given in Equation (3.2). If the plausible assumption is made that R_1 and R_2 are independent random variables, then Equation (3.3) can be simplified to

$$E(T_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T_1(T(r_1, r_2)) f_{R_1}(r_1) f_{R_2}(r_2) dr_1 dr_2 \quad (3.4)$$

where $f_{R_1}(r_1)$ and $f_{R_2}(r_2)$ are the probability density functions for rates R_1 and R_2 respectively. If we further choose to model R_1 and R_2 as independent uniform random variables uniformly distributed over the ranges $[R_{1,\min}, R_{1,\max}]$ and $[R_{2,\min}, R_{2,\max}]$ respectively, then Equation (3.4) simplifies further to

$$E(T_1) = \frac{1}{(R_{1,\max} - R_{1,\min})(R_{2,\max} - R_{2,\min})} \int_{R_{2,\min}}^{R_{2,\max}} \int_{R_{1,\min}}^{R_{1,\max}} T_1(T(r_1, r_2)) dr_1 dr_2. \quad (3.5)$$

Equation 3.5 is a complicated equation, and using manual computation would be an inefficient process. By contrast, numerical integration with the use of MATLAB can provide quick, comparable results. The MATLAB code for numerical analysis is contained in Appendix C.

For the initial scenario considered, both R_1 and R_2 were assumed to be uniform random variables over the range [1/hour, 10/hour], and both conditional fare values were identical, i.e., $I_1 = I_2 = \$100$. The known expected optimum time, $E(T_1)$, was predicted to be half of T_{total} because the two tasks have identical conditional fares, identical rates, and the time allotted cannot favor either task. Indeed, this test case did yield the known answer of $E(T_1) = 0.5T_{total}$, giving us initial confidence that our mathematics and methodology were correct.

With test results satisfactory, the next step was to create various ranges for uniform R_1 and R_2 . This thesis used ranges that were run multiple times and included small values, significantly larger values, and values of little variation; changing one variable in Equation (3.3) helps determine effects on the average optimum time of the asset. These sample sets were considered in order to mimic the random nature of the possible values that affect the optimum time. The smallest number in the range, R_{\min} , is the random variable minimum value, and the largest number in the range is the random variable maximum value, R_{\max} , to be substituted into Equation (3.3). Refer to Chapter IV for all values that are substituted for R_1 and R_2 , separately. Next, the focus was on I_1 with multiple different values while R_1 , R_2 , and I_2 remain consistent throughout the scenario. The last experiment observes how I_2 effects T_1 while R_1 , R_2 , and I_1 remain consistent throughout.

B. MONTE CARLO SIMULATION

This section illustrates how Monte Carlo simulation manipulates Equations (3.1) and (3.2) to simulate real-world events to find the average optimum time. Monte Carlo simulation is required to be mathematically independent from numerical analysis to verify the accuracy of the results. These results predict the optimum time allocation, which may increase the overall fare collected by the taxi or the overall intelligence value gathered by the asset. The Monte Carlo simulation also can be used to tell us more about the distribution of T_1 beyond just the expectation. The Monte Carlo simulation values can be used to form a histogram of T_1 values using the whole range of values for R_1 and R_2 , each. There are two reasons to use the Monte Carlo simulation as a second method. The first reason is to confirm what the numerical analysis finds for T_1 and the second argument is to provide more detail on the distribution of the optimum time allocations. Refer to Chapter IV for histograms.

In Monte Carlo simulation, numerical results emerge from repeatedly applying random variables to a computational algorithm [13]. Monte Carlo simulation can estimate

the expectation of random variable, T . In this method, the random variable T is assumed to be a function of other random variables X_1, X_2, \dots, X_N , i.e. $T = g(X_1, X_2, \dots, X_N)$. The joint probability density function of X_1, X_2, \dots, X_N is assumed to be known. First, a random N -tuple (x_1, x_2, \dots, x_N) is chosen consistent with the known joint probability density function. Second, the corresponding value for T is calculated using $t = g(x_1, x_2, \dots, x_N)$. These two steps are completed a large number of times, m , resulting in m values of t . These m values are averaged using the simplistic method of adding them and then dividing the sum by m . The resulting average is the Monte Carlo estimate of the expectation of T , herein called $MCE(T)$.

In this work, the Monte Carlo simulation was conducted as follows:

1. R_1 was chosen uniformly from the range $[R_{1,\min}, R_{1,\max}]$,
2. R_2 was chosen uniformly from the range $[R_{2,\min}, R_{2,\max}]$,
3. These values for R_1 and R_2 are substituted into Equations (3.1) and (3.2) yielding the corresponding value of T_1 .
4. Steps 1–3 were repeated $m-1$ more times.
5. The m values of T_1 are added and the sum is divided by m to yield the Monte Carlo estimate of $E(T_1)$, i.e., $MCE(T_1)$.

A computer was used for all five steps with $m = 10^6$. The MATLAB code for this Monte Carlo simulation is contained in Appendix D.

One of the key assumptions of our Monte Carlo simulation is that the number of trials, in this instance a million, is sufficient to make our estimate $MCE(T_1)$ to be very close to $E(T_1)$. The law of large numbers is used to increase the probability that the trials will replicate real-world events, known as relative frequency. The expected values of T_1 resulting from numerical analysis, $E(T_1)$, and Monte Carlo simulation, $MCE(T_1)$ must match very closely, assuming m is large. The uniform distribution was used to allow the results to be checked manually and quickly.

The Monte Carlo estimate of the expectation of T_1 will be compared to the $E(T_1)$ yielded by numerical analysis. First, our Monte Carlo algorithm was verified by running three simple tests separately in MATLAB and Excel. The two applications are necessary to provide verification of the $MCE(T_1)$ results and to make apparent any debugging in the MATLAB code. These tests consisted of verifying the average of random variable R_1 and random variable R_2 over a million times. The last simple test was averaging R_1 and R_2 jointly and comparing the results generated by MATLAB and Excel. Once the simple tests showed satisfactory results, the Monte Carlo code was run for the same sample sets that were used for R_1 , R_2 , I_1 , and I_2 in our numerical analysis method.

C. OVERVIEW

In summary, numerical analysis used the double integral in Equation (3.5) to determine the expected values for T_1 . Monte Carlo simulation was used to estimate the same $E(T_1)$ by generating a million values of T_1 consistent with the joint distribution of random variables R_1 and R_2 , and then averaging those million values. The $E(T_1)$ values from the numerical analysis and the $MCE(T_1)$ values from the Monte Carlo simulation will be compared in Chapter IV.

IV. RESULTS

This chapter discusses results for the expected optimum time allocation for Passenger One, i.e., $E(T_1)$ produced by numerical analysis and Monte Carlo simulation. The results are statistically identical, thus verifying the accuracy of the expected optimum time to pursue Passenger One, $E(T_1)$, values produced for each scenario of varying values for the hail rates for both passengers, R_1 and R_2 , and the conditional fare values of both passengers, I_1 , and I_2 . In the remainder of this thesis, the expected total fare collected, $E(\text{total fare collected})$, Equation (B.8) in Appendix B will be maximized with the expected optimum time to pursue both passengers, $E(T_1)$ and $E(T_2)$. The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$ Equation (B.4), will be used in an attempt to find a correlation between the expected optimum time to pursue both passengers, $E(T_1)$ and $E(T_2)$, values and parameters. Equations (B.4) and (B.8) proves that any modifications to the expected optimum time results, found using numerical analysis and Monte Carlo simulation, could cause the taxi to lose money and not gain the maximum amount of fare money that could be obtained in the available time duration.

A. COMPARISON OF RESULTS

This section scrutinizes 55 possible taxi scenarios resulting in the expected optimum time to pursue both passengers, $E(T_1)$ and $E(T_2)$, for each scenario. Numerical analysis and Monte Carlo simulation optimum expectation of time to pursue Passenger One, $E(T_1)$, columns match, allowing the optimum expectation of time to pursue Passenger Two, $E(T_2)$, to be calculated a single time for each scenario. The value for optimum expectation of time to pursue Passenger Two, $E(T_2)$, is easily found by subtracting the expected optimum time to pursue Passenger One, $E(T_1)$, from the total duration, also known as T_{total} . For each scenario, the expected optimum time allocation to

pursue each passenger, $E(T_1)$ and $E(T_2)$, for the taxi is calculated assuming the taxi is available for one hour. The optimum time allocation is defined to be the allocation that corresponds to the highest expected total fare during the one hour assumed available for the taxi. In the first group of scenarios, the scenarios are identical except for the distributions of the rate of hails for Passenger One, R_1 . In the second group, the scenarios are identical except for the distributions of the rate of hails for Passenger Two, R_2 . In the third group, the scenarios are identical expect for the distributions of the conditional fare of Passenger One, I_1 . In the fourth group, the scenarios are identical expect for the distributions of the conditional fare of Passenger Two, I_2 . In the fifth group, the scenarios are identical except the distributions of the conditional fare of Passenger One, I_1 with different but constant hail rates for both passengers, R_1 and R_2 . The sixth group of scenarios are identical expect the distributions of the conditional fare of Passenger One, I_1 , with a small hail rate for Passenger One, R_1 , compared to the hail rate for Passenger Two, R_2 . The last group of scenarios are identical expect for the distributions of the conditional fare of Passenger One, I_1 , with a much smaller hail rate for Passenger One, R_1 , compared to the hail rate for Passenger Two, R_2 .

1. Group One

The results of group one, scenarios one through eight, for the expected optimum time to pursue Passenger One, $E(T_1)$, obtained were identical to a fraction of a percentage for all scenarios performed for various ranges of uniformly distributed hail rates of Passenger One, R_1 , as displayed in Table 1. The final results of the expected optimum time to pursue Passenger One, $E(T_1)$, is explained by calculations found in Appendix B for the probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, and the total conditional expected fare value, $E(\text{total fare collected})$, and calculations found in Appendix E to find the probability of the hail rate of Passenger One, R_1 , being larger than the hail rate of Passenger Two, R_2 ,

$\Pr(R_1 > R_2)$. The mentioned probabilities of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, and the hail rate of Passenger One, R_1 , being larger than the hail rate of Passenger Two, R_2 , will be used to support why the taxi time allocation results are as they are. The total conditional expected fare value, $E(\text{total fare collected})$, is examined to explain what the taxi driver gains from pursuing that passenger for the recommended time compared to other total expected fare values that are not at the expected optimum times, $E(T_1)$ and $E(T_2)$.

The expected optimum time to pursue Passenger One, $E(T_1)$, found in Table 1 is the value for which the expected total fare collected, $E(\text{total fare collected})$, is a maximum value. If a time, Δt , is stolen from Passenger One and given to Passenger Two, making $T_1 < E(T_1)$, then the expected fare from Passenger One, $E(\text{fare from Passenger One})$, must drop in value while the expected fare from Passenger Two, $E(\text{fare from Passenger Two})$ will increase. Clearly, the increase in the expected fare from Passenger Two, $E(\text{fare from Passenger Two})$ is more than compensated for by the decrease in the expected fare from Passenger One, $E(\text{fare from Passenger One})$. This is true here and for all scenarios.

Table 1. Results for numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various uniform distributions for the hail rate for Passenger One, R_1 .

Minimum of $R_2 = [1/\text{hour}]$, Maximum of $R_2 = [10/\text{hour}]$, $I_1 = \$100$, $I_2 = \$100$, $T_{total} = 60$ minutes					
Scenario	Minimum of R_1 (/ hour)	Maximum of R_1 (/ hour)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
1	1.00	10.00	30.00	30.00	30.00
2	1.00	25.00	22.37	22.37	37.63
3	10.00	1×10^6	0.01	0.01	59.99
4	1.00	2.00	35.51	35.51	24.49
5	1.00	5.00	33.96	33.96	26.04
6	1.00	7.00	32.32	32.32	27.68
7	0.11	0.25	18.66	18.66	41.34
8	0.10	0.101	14.61	14.62	45.38

a. Scenario One

The expected optimum values of time to pursue Passenger One, $E(T_1)$, in Table 1 are applicable to the previously mentioned taxi scenario. Recall that the focus is to find the optimum time to allocate the limited resource, the taxi, to pursuing the conditional fares of Passenger One and Passenger Two. For example, if prior experience shows that the hail rate for Passenger One, R_1 , can vary between 1 hail per hour and 10 hails per hour and the rate for Passenger Two, R_2 , is also between 1 hail per hour and 10 hails per hour, and the conditional fare value of both tasks is $I_1 = I_2 = \$100$, then the expected optimum time allocated to pursuing Passenger One and Passenger Two is exactly 30 minutes each. Figure 3 is the histogram from the Monte Carlo simulation for Scenario One. This scenario was considered as a check of the algorithm. Since each passenger has the same conditional fare value and the same hail rate, then it makes sense that the optimum solution would not favor either passenger, since they are indistinguishable in the parameters that impact the expected total fare value. Therefore, we know the optimum solution is to split the hour evenly between the two tasks. Furthermore, the rates

are just as likely to favor one passenger as the other; therefore, the histograms are symmetrical. The probability density functions for optimum allocation times, T_1 and T_2 , were not calculated in this work. However, histograms have approximately the same shape as the corresponding probability density functions. Therefore, optimum allocation times T_1 and T_2 have probability density functions with the same shapes as the histograms in Figure 3. Notice that not only does the Monte Carlo simulation confirm the expected optimum time allocation is 30 minutes for each, but it shows that the optimum time of both passengers, T_1 and T_2 , have the same, symmetrical, probability distributions. These histograms reveal that under no circumstances consistent with Scenario One, can we expect to optimize the fare if less than 18 minutes (i.e., 0.3 hours,) is allocated to seeking either passenger. That is useful information for the taxi driver seeking to maximize his fares for the assumed available hour. The histogram of the optimum time to pursue Passenger Two, T_2 , is a mirror image and is not shown for other scenarios.

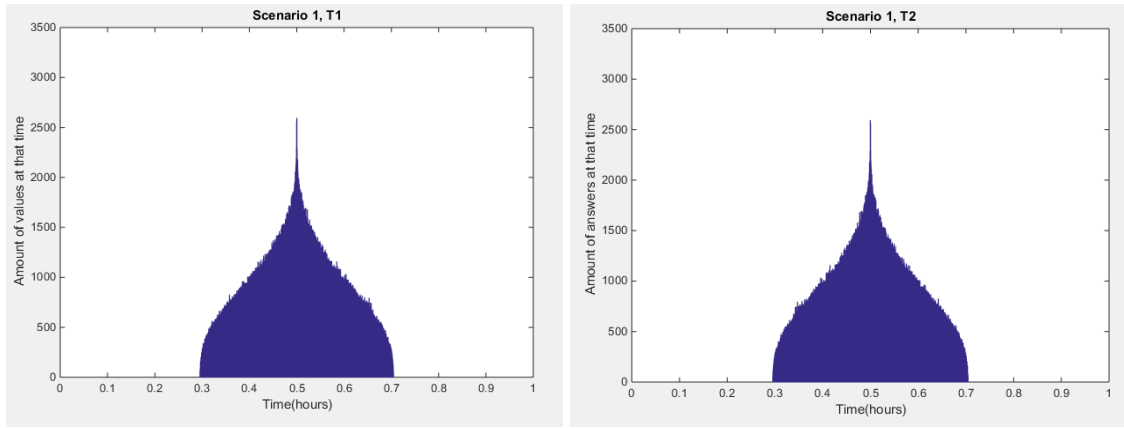


Figure 3. Scenario One histogram.

b. Scenario Two and Three

Examining Scenario Two reveals the expected optimum time for a hail rate that is potentially larger means less time spent pursuing that passenger under these circumstances. The circumstances of Scenario Two are the hail rate for Passenger One, R_1 , can vary between 1 hail per hour and 25 hails per hour and the hail rate for Passenger

Two, R_2 , is between 1 hail per hour and 10 hails per hour, and the conditional fare values of both tasks, I_1 and I_2 , remain at the \$100 value, then the expected optimum time to pursue Passenger One, $E(T_1)$, is 22.4 minutes and the expected optimum time to pursue Passenger Two, $E(T_2)$, is 37.6 minutes. This scenario allows more time to be spent seeking Passenger Two due to the larger probability of successfully pursuing Passenger One. $\Pr(\text{Passenger One is picked up})$ maximum value is nearly one (for the hail rate of Passenger One, R_1 , between 10 hails per hour and 25 hails per hour). The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is calculated in Appendix B Equation (B.4) and proves that a potentially quicker hail rate is an influential parameter for obtaining the conditional fare money. The results of Scenario Two, in Table 1, make sense because the probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , i.e. $\Pr(R_1 > R_2)$, is equal to 0.8125. This is caused by Passenger One usually hailing faster due to the hail rate of Passenger One, R_1 , being between 11 hails per hour and 25 hails per hour. The probability calculation of $\Pr(R_1 > R_2)$ for Scenario Two is found in Appendix E.

Figure 4 is the histogram from the Monte Carlo simulation of Scenario Two that displays the mode of Scenario Two to be approximately 18 minutes. For this scenario, the recommendation to the taxi driver is the expected optimum time to achieve the taxi fare of \$100, if the passenger is successfully picked up, is approximately 22.4 minutes. Observe that the expected optimum time to pursue Passenger One, $E(T_1)$, in Scenario Two in Table 1 is not the same as the peak of the histogram in Figure 4.

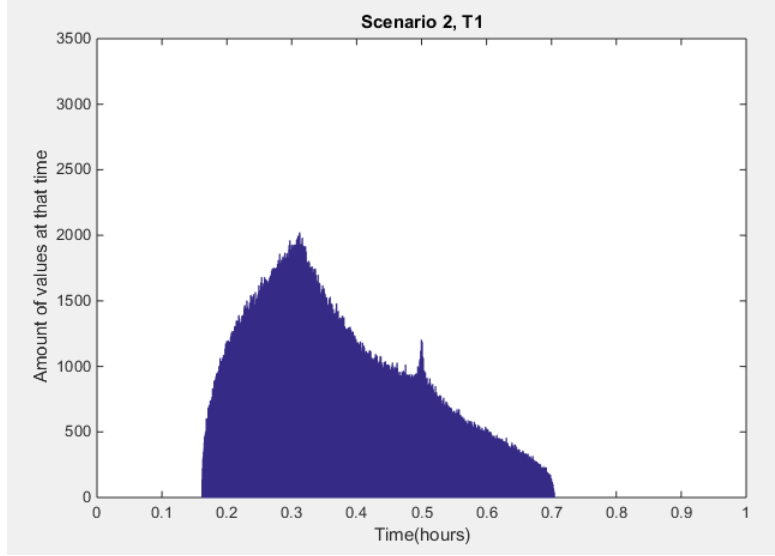


Figure 4. Scenario Two histogram.

This is an example of the relationship between the mean, also known as the expected optimum time to pursue Passenger One, $E(T_1)$, and the mode which is displayed as the highest peak in Figure 4 and is the most likely value for T_1 . This explains why the results in Table 1 do not align with Figure 4 peaks and crests. Figure 4 indicates that pursuing Passenger One for approximately 10 minutes to 42 minutes will yield \$100 in fare money compared to pursuing Passenger One for the average time to successfully pursue is 22.4 minutes. The expected optimum time to pursue Passenger One is 22.4 minutes to maximize the expected overall fare collected, $E(\text{total fare collected})$.

Scenario three reveals the expected optimum time for a hail rate that is much larger means a small amount of time will be recommended to pursue that passenger. The circumstances of Scenario three are the hail rate for Passenger One, R_1 , can vary between 10 hails per hour and 1,000,000 hails per hour and the hail rate for Passenger Two is between 1 hail per hour and 10 hails per hour, with the conditional fare of each task remains \$100, then the expected optimum time to pursue Passenger One, $E(T_1)$, is 0.1 minute and the expected optimum time to pursue Passenger Two, $E(T_2)$, is 59.9 minutes. This scenario allows more time to be spent seeking Passenger Two due to

the larger probability, of successfully pursuing Passenger One. $\Pr(\text{Passenger One is picked up})$ maxing out at nearly one, using Equation (B.4). The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, proves that the quicker hail rate is an influential parameter. The results of Scenario three, in Table 1, make sense because the probability the hail rate of Passenger One, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$ is equal to one. This is due to the slowest possible hail for Passenger One (10 hails per hour) being larger than the fastest possible hail for Passenger Two (10 hails per hour). The probability calculation of $\Pr(R_1 > R_2)$ for Scenario three uses Equation (E.2).

c. Scenarios Seven and Eight

Notably, a majority of the expected optimum time to pursue Passenger One, $E(T_1)$, values in Table 1 are less than 30 minutes regardless of Passenger One being found faster (Scenarios two and three) or slower (Scenarios seven and eight). There are two reasons for the expected optimum time to pursue Passenger One, $E(T_1)$, to be less than 30 minutes. One reason is the scenarios' parameters ensure more hails for Passenger One, R_1 , in less time when compared to Passenger Two hail rate parameters, R_2 , as explained in section IV.A.1.b, thus less than half of the taxi time is allocated to $E(T_1)$ and the other reason is that Scenarios seven and eight involve a smaller probability of Passenger One being picked up, $\Pr(\text{Passenger One is picked up})$ Equation (B.4), thus leaving more time for Passenger Two to be pursued for the same conditional fare value, I_2 . Scenario seven's probability of Passenger One being successfully picked up, $\Pr(\text{Passenger One is picked up})$, for such circumstances range from 0.034 to 0.075. Scenario eight's probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is 0.02. The small probabilities of Passenger One being picked up, $\Pr(\text{Passenger One is picked up})$, and the small conditional expected fare for Passenger One, $E(\text{fare from Passenger One})$ Equation (B.1), are the reason that the expected optimum time pursuing Passenger Two, $E(T_2)$, are so large for Scenarios seven

and eight. Scenarios seven and eight probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$ is equal to zero, using Equation (E.2). This is due to Passenger One hail rate being much smaller than Passenger Two's hail rate. Figure 5 is the histogram of Scenario seven.

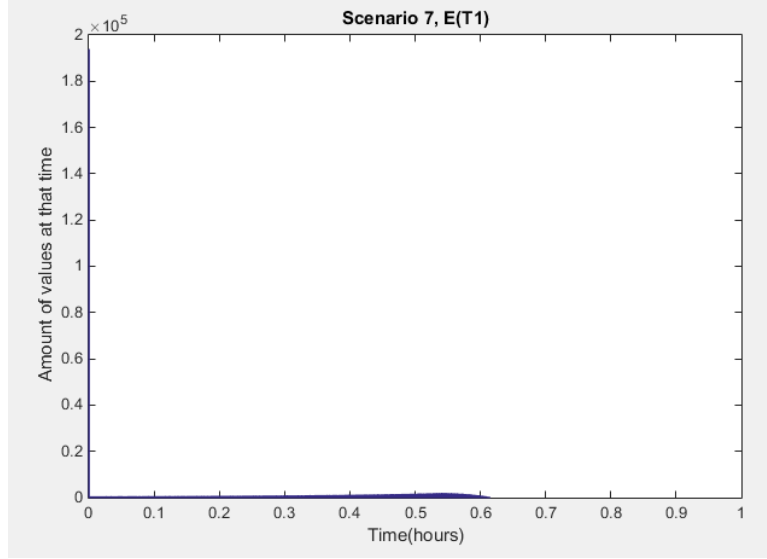


Figure 5. Scenario Seven histogram.

The histogram of Scenario seven, in Figure 5, shows these circumstance are likely to allocate zero time to T_1 , peak of Figure 5, however the expected optimum time to pursue Passenger One, $E(T_1)$, is 18.6 minutes (0.31 of an hour). This is another example of where the mode of T_1 and the expected optimum time pursuing Passenger One, $E(T_1)$, does not align. Notice the larger y-axis scale of Figure 5 compared to previous figures. This parameters reflect the low probability of Passenger One being picked up, $\Pr(\text{Passenger One is picked up})$, and the conditional fare value not compensating for this low probability.

d. Scenarios Four through Six

The last examination of Table 1 focuses on Scenarios four through six. The expected optimum time pursuing Passenger One, $E(T_1)$, values recommended are more

than half the assumed available hour, T_{total} . This is anticipated by performing similar calculations found in Appendix B for the probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, and the probability of the hail rate of Passenger One, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$, found in Appendix E. The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, for Scenario four is 0.45 to 0.69. The probability the hail rate of Passenger One, R_1 , is larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$ equals 0.44, for Scenario four. For Scenario five the probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is 0.43 to 0.94 while the probability of the hail rate of Passenger One, R_1 , is larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$, is 0.28. Scenario six circumstances probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, are 0.42 to 0.97 while the probability of the hail rate of Passenger One, R_1 , is larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$, is 0.17. This was to be predicted as the hail rates for Passenger One, R_1 , are smaller than the hail rates of Passenger Two, R_2 , and thus requiring more taxi time to accomplish pursuing Passenger One, $E(T_1)$, but as the range of hail rates for Passenger One, R_1 becomes more wide spread the probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, increases from Scenario four to Scenario seven.

Scenarios four through six reveal that more time should be spent seeking Passenger One than Passenger Two, largely due to the hail rate of Passenger One, R_1 , being smaller than the hail rate of Passenger Two, R_2 , for the same conditional fare value. The T_1 for Scenario four is displayed in Figure 6.

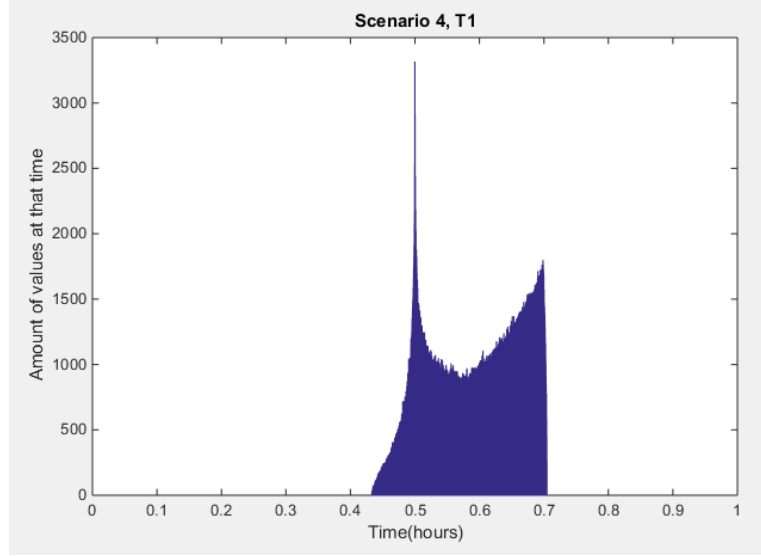


Figure 6. Scenario Four histogram.

2. Group Two

The next set of scenarios focuses on varying the distribution for uniform random variable hail rate for Passenger Two, R_2 . These scenarios were run to fully explore the relationship between the hail rate for Passenger Two, R_2 , and the expected optimum time to pursue Passenger One, $E(T_1)$. Table 2 displays the values of the expected optimum time pursuing Passenger One, $E(T_1)$. The key observations from Table 2 are that the results from numerical analysis and Monte Carlo simulation match, and when comparing Table 1 and Table 2, the expected optimum time to pursue the passengers, $E(T_1)$ and $E(T_2)$, values are interchanged. This was expected as the parameters for the hail rates of the passengers, R_1 and R_2 , are reversed thus resulting in expected optimum time to pursue the passengers, $E(T_1)$ and $E(T_2)$, values also being reversed. Comparing Figure 4 and the histogram of Scenario 10, in Figure 7 concurs with this observation. The observations from section IV.A.1 apply to this group of scenarios.

Table 2. Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various uniform distributions of hail rate for Passenger Two, R_2 .

Minimum of $R_1 = [1/\text{hour}]$, Maximum of $R_1 = [10/\text{hour}]$, $I_1 = \$100$, $I_2 = \$100$, $T_{total} = 1 \text{ hour}$					
Scenario	Minimum of R_2 (/ hour)	Maximum of R_2 (/ hour)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
9	1.00	10.00	30.00	30.00	30.00
10	1.00	25.00	37.63	37.62	22.38
11	10.00	1×10^6	59.99	59.99	0.01
12	1.00	2.00	24.49	24.50	35.50
13	1.00	5.00	26.04	26.04	33.96
14	1.00	7.00	27.68	27.69	32.31
15	0.11	0.25	41.33	41.34	18.66
16	0.10	0.101	45.39	45.41	14.59

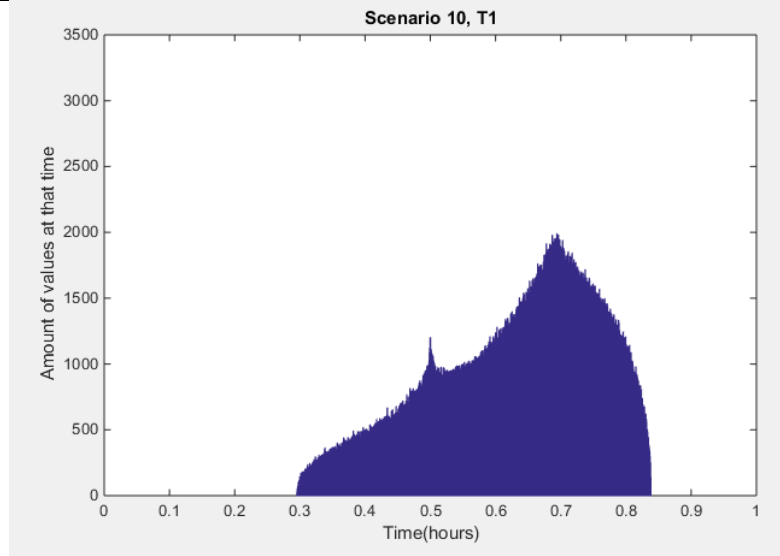


Figure 7. Scenario 10 histogram.

3. Group Three and Group Four

The next set of scenarios examines the expected optimum time pursuing Passenger One, $E(T_1)$, with respect to various conditional fare values for the passengers,

I_1 and I_2 . Table 3 displays the results of the expected optimum time to pursue Passenger One, $E(T_1)$, with changing conditional fare values for Passenger One, I_1 , values, while Table 4 displays the results for the expected optimum time to pursue Passenger One, $E(T_1)$, with changing conditional fare values for Passenger Two, I_2 , values. Indeed, Table 3 and Table 4 document the same hail rates for both passengers, R_1 and R_2 . Not surprisingly, the values of expected optimum time to pursue the passengers, $E(T_1)$ and $E(T_2)$, are interchanged when comparing Table 3 and Table 4. Due to this simple relationship observed between the scenarios in Table 3 and Table 4, only observations about the scenarios in Table 3 will be discussed in detail. The stated observations for Table 3 scenarios apply to the Table 4 scenarios, but with Passengers One and Two exchanged. Scenario 17 was used as a check for the algorithm as seen in Scenarios one and nine. The histogram for Scenario 17 is the same as Figure 3.

Table 3. Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various conditional fare values for Passenger One, I_1 .

Minimum of $R_1 = [1/\text{hour}]$, Maximum of $R_1 = [10/\text{hour}]$, minimum of $R_2 = [1/\text{hour}]$, maximum of $R_2 = [10/\text{hour}]$, $I_2 = \$100$, $T_{total} = 1$ hour					
Scenario	I_1 (\$)	I_2 (\$)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
17	100.00	100.00	30.00	30.00	30.00
18	101.00	100.00	30.06	30.05	29.95
19	1×10^3	100.00	44.24	44.24	15.76
20	1×10^4	100.00	53.90	53.92	6.08
21	1×10^5	100.00	58.56	58.57	1.43
22	99.00	100.00	29.93	29.95	30.05
23	50.00	100.00	25.61	25.61	34.39
24	25.00	100.00	21.22	21.24	38.76

a. Scenario 18

Scenario 18 includes the conditional fare value for Passenger One, I_1 , to be \$101 while the conditional fare values for Passenger Two, I_2 , was \$100, and the hail rates for the taxi are still uniformly varying from 1 hail per hour to 10 hails per hour for both passengers, R_1 and R_2 . Then the resultant expected optimum time pursuing Passenger One, $E(T_1)$, is 30.1 minutes. The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is equal to a minimum of 0.39 to a maximum of 0.99 and is the same for Passenger Two because the hail rate parameters, R_1 and R_2 , are identical. This example illustrates that the expected optimum time to pursue Passenger One, $E(T_1)$, increases when the conditional fare value for Passenger One, I_1 , is increased. The increase of the conditional fare value by \$1 equates to approximately half a minute of taxi time. This statement comes from comparing Scenario 18 to Scenario 17. The T_1 for Scenario 18 is displayed in Figure 8.

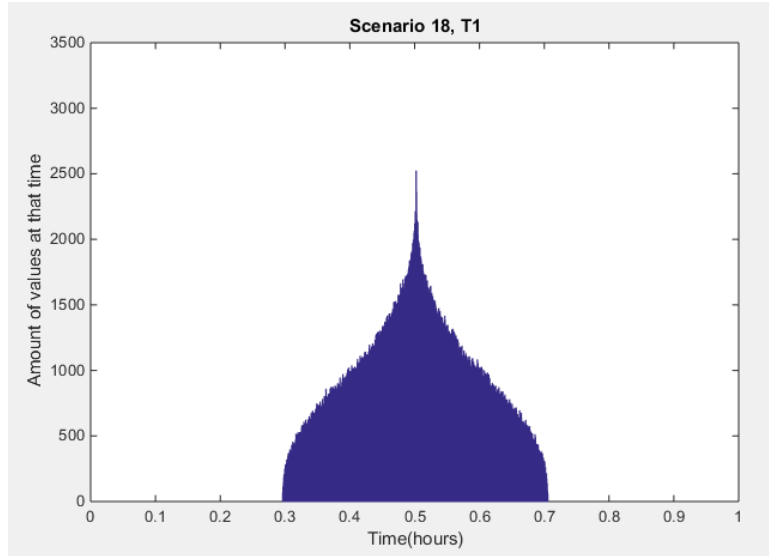


Figure 8. Scenario 18 histogram.

Scenarios 18 through 21 are all cases where the conditional fare value of Passenger One, I_1 , was larger than the conditional fare value of Passenger Two, I_2 , and

since the hail rates of both passengers are identical all the results show a preference, of varying degrees, for securing the conditional fare that is larger, i.e., Passenger One's, I_1 . Scenarios 22 through 24 further support that the algorithms' preference for allocating more time to secure the conditional fare value that is larger under these circumstances.

Table 4. Results for Numerical analysis (Equation (3.5)) and Monte Carlo simulation averaging to determine the expected optimum time pursuing Passenger One, $E(T_1)$, with various conditional fare values for Passenger Two, I_2 .

Minimum of $R_1 = [1/\text{hour}]$, Maximum of $R_1 = [10/\text{hour}]$, minimum of $R_2 = [1/\text{hour}]$, maximum of $R_2 = [10/\text{hour}]$, $I_1 = \$100$, $T_{\text{total}} = 1 \text{ hour}$					
Scenario	I_1 (\$)	I_2 (\$)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
25	100.00	100.00	30.00	29.99	30.01
26	100.00	101.00	29.94	29.94	30.06
27	100.00	1×10^3	15.76	15.79	44.21
28	100.00	1×10^4	6.10	6.10	53.90
29	100.00	1×10^5	1.44	1.44	58.56
30	100.00	99.00	30.07	30.07	29.93
31	100.00	50.00	34.39	34.39	25.61
32	100.00	25.00	38.76	38.76	21.24

4. Group Five

Group Five scenarios examine the effect of different hail rates between Passenger One and Two, R_1 and R_2 , with varying conditional fare values for Passenger One, I_1 . Different combinations of hail rates with conditional fare values must be explored to resemble possible circumstances that the taxi driver may encounter. Studying the combinations of different hail rates between Passenger One and Two, R_1 and R_2 , with varying conditional fares for Passenger One, I_1 , shows a readily apparent effect of one of the parameters on the expected optimum time of Passenger One, $E(T_1)$. If the taxi driver

only considered the fare value to base his decision on how to allocation his time to pursuing passengers, then he would be completely ignoring the probability of collection which takes into account the rates of hailing by that passenger. This is a less then optimized method to determine to time allocation and illustrates the importance the hail rate have in determining the optimized expected total fare collected. The relationship between conditional fare values and time allocated is complex as shown by Equation (1.2) but also must consider the hail rates. All the scenarios in this group have a probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$ equal to 0.81.

Table 5. Resultant expected optimum time for Passenger One, $E(T_1)$, for various conditional fare values for Passenger One, I_1 , and different hail rate for Passenger One, R_1 , and Passenger Two, R_2 .

Minimum of $R_2 = [1/\text{hour}]$, Maximum of $R_2 = [10/\text{hour}]$, $I_2 = \$100$, $T_{total} = 1$ hour						
Scenario	Minimum of R_1 (/hour)	Maximum of R_1 (/hour)	I_1 (\$)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
33	1.00	25.00	101.0	22.42	22.42	37.58
34	1.00	25.00	1×10^3	31.67	31.66	28.34
35	1.00	25.00	1×10^4	39.25	39.22	20.78
36	1.00	25.00	1×10^5	44.96	44.95	15.05
37	1.00	25.00	99.00	22.33	22.33	37.67
38	1.00	25.00	50.00	19.54	19.54	40.46
39	1.00	25.00	25.00	16.70	16.72	43.28

a. Table 5 Compared to Table 3

Comparing Table 5 and Table 3, it becomes apparent that the expected optimum time to pursue Passenger One, $E(T_1)$, for Scenarios 33 through 36 are decreased compared to Scenarios 18 through 21. The resultant expected optimum time pursuing Passenger One, $E(T_1)$, is affected by the wide range hail rate for Passenger One, R_1 as

seen in circumstances of Scenarios 33 through 36. The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, for circumstances of Scenarios 33 to 36 range widely but all have a maximum value of almost one. This is interesting because comparing Table 3 Scenarios 18 through 21 to Table 5 Scenarios 33 through 36, Table 3 probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, results in the maximum value of nearly one for only Scenarios 20 and 21 whereas the Scenarios 33 through 36 probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is nearly one for all scenarios. This is evidence that the hail rates effect the expected optimum time to pursue Passenger One, $E(T_1)$ and must be considered to accurately optimize time allocation of the taxi. For example if the taxi driver only considered the conditional fare values, I_1 and I_2 , to allocate available time, like in Table 3, then (s)he would have wasted time by over allocating approximately 12 to 14 minutes for circumstances of Scenarios 34 through 36 in comparison with Scenarios 19 through 21. This same comparison can be made for Scenarios 33 and 37 through 39 to Scenarios 18 and 22 through 24. The expected optimum times, $E(T_1)$ and $E(T_2)$, difference between the groups of scenarios is five minutes to eight minutes. This is an easily apparent example of how optimizing the taxi time allocation is effected by the hail rates of the passengers, R_1 and R_2 .

b. Scenario 36

Scenario 36 is a case where the hail rate of Passenger One, R_1 , and the conditional fare value of Passenger One, I_1 , is much higher than both the hail rates of Passenger Two, R_2 , and the conditional fare value of Passenger Two, I_2 . Notice that Passenger One has a much larger expected optimum time allocation for Passenger One, $E(T_1)$, compared to Passenger Two, $E(T_2)$ but not all of the taxi time is spent on Passenger One. This scenario is an example of where the probability of picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is greater than 0.50 for all possible hail rates of Passenger One and allows the taxi time to be allocated to Passenger Two to maximize the

total conditional expected fare value, $E(\text{total fare collected})$. This scenario also contradicts scenario three observation of the larger hail rate equates to a smaller value for the expected optimum time to pursue Passenger One, $E(T_1)$. Scenario 36 exhibits circumstances where the conditional fare value for Passenger One, I_1 , compensates for the hail rate of Passenger One, R_1 .

5. Group Six

The following scenarios examine how the relationship between a smaller hail rate of Passenger One, R_1 , and varying conditional fare values for Passenger One, I_1 , increases the expected optimum time spent pursuing Passenger One, $E(T_1)$, when compared to Table 5. Table 6 displays the expected optimum time pursuing Passenger One, $E(T_1)$, values for Scenarios 40 through 47. All scenarios in this group have a probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , i.e., $\Pr(R_1 > R_2) = 0.44$ based on the hail rate parameters of these scenarios.

Table 6. Resultant expected optimum time pursuing Passenger One, $E(T_1)$ for various conditional fare values for Passenger One, I_1 and $\Pr(R_1 > R_2) < 0.5$.

Minimum of $R_2 = [1/\text{hour}]$, Maximum of $R_2 = [10/\text{hour}]$, $I_2 = \$100$, $T_{\text{total}} = 1 \text{ hour}$						
Scenario	Minimum of R_1 (/hour)	Maximum of R_1 (/hour)	I_1 (\$)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
40	1.00	2.00	101.0	35.61	35.61	24.39
41	1.00	2.00	1×10^3	56.62	56.61	3.39
42	1.00	2.00	1×10^4	60.00	60.00	0.00
43	1.00	2.00	1×10^5	60.00	60.00	0.00
44	1.00	2.00	99.00	35.41	35.41	24.59
45	1.00	2.00	58.50	29.98	29.99	30.01
46	1.00	2.00	50.00	28.43	28.42	31.58
47	1.00	2.00	25.00	21.47	21.48	38.52

a. *Effects of a Close Range of Hail Rates*

Examination of Table 6 reveals that the hail rate of Passenger One, R_1 , has a high probability of being smaller than the hail rate of Passenger Two, R_2 , and the expected optimum time spent pursuing Passenger One, $E(T_1)$ will be larger. The higher conditional fare values for Passenger One, I_1 , as seen in Scenarios 40 through 43 are compensation for the low hail rates of Passenger One, R_1 , and recommend the time allocation of the taxi be more than half of the total time available, T_{total} . Scenario 44 recommends the taxi driver spend more than 30 minutes pursuing Passenger One, $E(T_1)$, regardless of the smaller conditional fare value of Passenger One, I_1 . As can be seen from Scenarios 46 and 47, the low hail rate of Passenger One, R_1 , combined with the low conditional fare value of Passenger One, I_1 , creates circumstances for Passenger Two parameters to have more of the taxi time allocation.

b. *Scenario 45*

Further study of these parameters, using numerical analysis and Monte Carlo simulation, revealed that the conditional fare value of Passenger One, I_1 , could be \$58.50 compared to the conditional fare value of Passenger Two, I_2 , equal to \$100 before the expected optimum time pursuing Passenger One, $E(T_1)$, and the expected optimum time pursuing Passenger Two, $E(T_2)$, were allocated equal amounts time. Figure 9 displays the histogram of this unique scenario. This is an example of how the hail rates combined with the conditional fare values could affect the management of the taxi's time and shows the error in thinking that only the conditional fare value is the only parameter to be considered when trying to make the most money with the taxi.

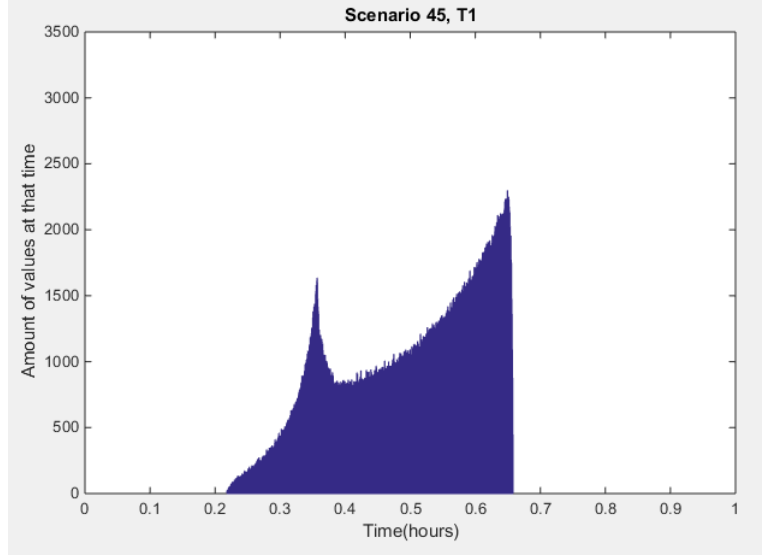


Figure 9. Scenario 45 histogram.

6. Group Seven

The last set of scenarios study the expected optimum time pursuing Passenger One, $E(T_1)$, with parameters that include hail rates of Passenger One, R_1 , to be significantly smaller, i.e., 0.11 hails per hour to 0.25 hails per hour (once every four to nine hours), than the hail rates of Passenger Two, R_2 , which remain between 1 hail per hour and 10 hails per hour combined with varying values of the expected conditional fare for Passenger One, I_1 . Comparing Table 6 and Table 7 reveals how a significantly smaller hail rate of Passenger One, R_1 , affects the expected optimum time pursuing Passenger One. In many scenarios significant time is allocated to Passenger One, in spite of having a low hail rate. This is another illustration of the complex interaction of the two rates and the two conditional values with the necessity to consider all four factors to achieve optimum overall results. All scenarios in this group have a probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2)$ equal to one.

Table 7. Resultant expected optimum time pursuing Passenger One, $E(T_1)$ for various conditional fare values for Passenger One, I_1 , combined with a much smaller hail rate for Passenger One, R_1 .

Minimum of $R_2 = [1/\text{hour}]$, Maximum of $R_2 = [10/\text{hour}]$, $I_2 = \$100$, $T_{total} = 1$ hour						
Scenario	Minimum of R_1 (/hour)	Maximum of R_1 (/hour)	I_1 (\$)	Numerical Analysis $E(T_1)$ (minutes)	Monte Carlo simulation $E(T_1)$ (minutes)	$E(T_2)$ (minutes)
48	0.11	0.25	101.00	18.75	18.76	41.24
49	0.11	0.25	275.00	30.01	30.00	29.99
50	0.11	0.25	1×10^3	48.01	48.01	11.99
51	0.11	0.25	1×10^4	60.00	60.00	0.00
52	0.11	0.25	1×10^5	60.00	60.00	0.00
53	0.11	0.25	99.00	18.58	18.57	41.43
54	0.11	0.25	50.00	13.62	13.62	46.38
55	0.11	0.25	25.00	9.72	9.71	50.29

a. Scenario 51 and 52

Scenarios 51 and 52 recommend all the taxi time to pursuing Passenger One to optimize the total expected fare. The probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, for Scenarios 51 and 52 range from 0.10 to 0.22. The only reason that the algorithms would allocate all the taxi time to pursuing Passenger One, $E(T_1)$, to a passenger that hails at a rate less than once every four hours, is that the conditional fare value is significantly larger than Passenger Two's conditional fare value. This is an important discovery because the goal is to find the expected optimum time allocation to maximize the overall expected fare gathered. Amazingly, the algorithms recommend the taxi to pursue a passenger that only hails at most once every four hours for a conditional fare of \$10,000 or \$100,000 compared to a passenger that will hail once every hour at the least for \$100 fare. The other passenger, Passenger Two, has a probability of picking up Passenger Two, $\Pr(\text{Passenger Two is picked up})$, close to one but the small conditional fare of \$100 voids this high probability.

b. Table 7 versus Table 6

Comparing Table 6 and Table 7 reveals that a high probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, is not a requirement to have taxi time allocated. In Scenarios 48 through 55, the probability of successfully picking up Passenger One, $\Pr(\text{Passenger One is picked up})$, ranges from 0.02 to 0.22 and yet Scenarios 51 and 52 have all the taxi time allocated to pursuing Passenger One. This is what happened in Scenarios 42 and 43 but the hail rate of Passenger One, R_1 , in Scenarios 42 and 43 ranges from 1 hail per hour to 2 hails per hour. The probability of picking up a passenger does not account for the conditional fare value and is the reason why the taxi cannot solely base time allocation on the hail rates for passengers. The taxi driver must consider both the hail rates and fare values to optimize the overall money collected.

c. Scenario 49

In Scenario 49 the algorithms recommend the available time allocation to be evenly split between both passengers, $E(T_1)$ and $E(T_2)$. Scenario 49 reveals the conditional fare value of Passenger One, I_1 , must equal \$275 compared to the conditional fare value of Passenger Two, I_2 , equal to \$100 with the hail rates of the circumstance of Group Seven to allocate 30 minutes to Passenger One. This scenario illustrates how much compensation by the conditional fare value is needed to overcome the low hail rate of Passenger One, R_1 . Figure 10 displays the T_1 of Scenario 49.

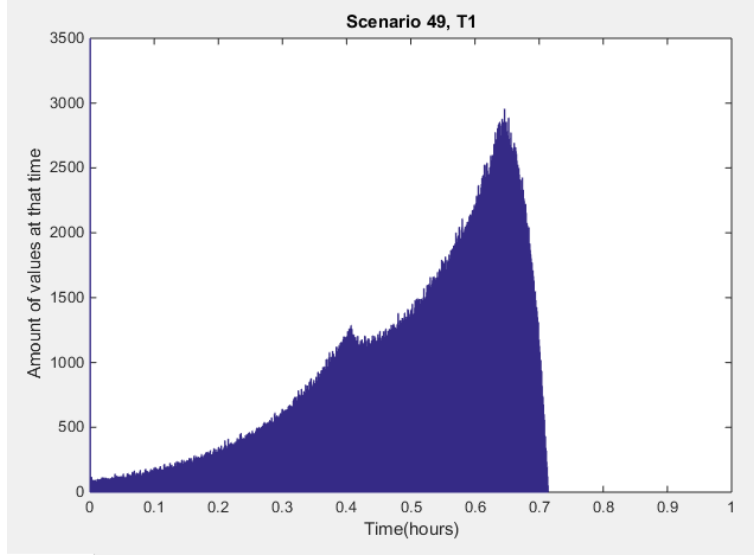


Figure 10. Scenario 49 histogram.

B. OVERVIEW

This chapter displays how the expected optimum time to pursue Passenger One, $E(T_1)$, varied for various uniform distributions of the hail rates, R_1 and R_2 , and various conditional fare values, I_1 and I_2 . It then focuses on illustrating the effects of the different hail rate ranges while varying one conditional fare value. The most significant revelations came from scenarios 36 (Table 5), 42, 43, 45 (Table 6), 49, 51, and 52 (Table 7). Scenario 36 shows that the probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2) = 0.8125$ combined with a much larger conditional fare value still allocates taxi time to the other passenger, Passenger Two, regardless of the smaller probability of successfully securing the taxi hail and fare. This scenario recommends the taxi driver should spend 45 minutes pursuing Passenger One with those circumstances of Scenario 36. This is due to the chance that Passenger One hail rate, R_1 , is less than Passenger Two in the assumed one hour available and the conditional fare value of Passenger One being \$100,000. Scenarios 42 and 43 include a probability of Passenger One hail rate, R_1 , being larger than the hail rate of Passenger Two, R_2 , $\Pr(R_1 > R_2) = 0.44$ and a conditional fare value of Passenger

One, I_1 , is two and three times more than the Passenger Two conditional fare value, I_2 . It is then recommended that all of the taxi time is allocated to pursue Passenger One. Scenarios 42 and 43 indicate that the conditional fare value is important to optimize the use of the taxi but the hail rates and the probabilities of success must be considered to fully understand how to optimize the taxi in the one hour available. Scenario 45 acknowledges the hail rate of Passenger One, R_1 , can be compensated with the conditional fare value to have split even time allocations between the passengers. The same revelation of Scenario 45 can be said for the Scenario 49. The conditional fare value of Passenger One, I_1 , in Scenario 49 is \$275 to compensate for a much smaller hail rate of Passenger One, R_1 , equal to once every four to nine hours, competing for the taxi time against Passenger Two with a large hail rate and small conditional fare value. The examination of Scenarios 51 and 52 reveals all the taxi time should be spent on Passenger One with a smaller probability of Passenger One hail rate, R_1 , than in Scenarios 42 and 43. Again the algorithms recommend all the taxi time to be spent on the higher conditional fare value passenger, Passenger One, regardless of the probability of the taxi being hailed by that passenger. This is due to the small $\Pr(\text{Passenger One is picked up})$ being more than sufficiently compensated by the conditional fare value of Passenger One, I_1 , if it is \$10,000 or more. Again illustrating the importance of considering all factors, hail rates and conditional fare values, to maximize the expected optimal total fare gathered. Chapter V discusses conclusions and suggests further areas for investigation.

V. SUMMARY OF FINDINGS AND AVENUES FOR FURTHER RESEARCH

This thesis proves that a complex relationship between the hail rates of both passengers, R_1 and R_2 , and the conditional fare values of both passengers, I_1 and I_2 , contribute to the dimensions that are necessary to find the expected optimum time of the taxi. Furthermore, it was shown that this ability does not rely on knowing parameters, such as the average rate of passengers hailing taxi, to any great precision. Allocating taxi time according to the expected optimum time of both passengers, $E(T_1)$ and $E(T_2)$, using numerical analysis and Monte Carlo simulation to the taxi scenario means the total fare gathered in the assumed one hour of availability is maximized. This was proven to be true in Chapter IV. However, there are still unexplored avenues to consider. Unexplored considerations to be investigated include considering the conditional fare values as random variables in the same fashion as the rates or finding the optimum allocation time for more than two tasks during the assumed available time of the taxi.

A. CONCLUSION

The examination of scenarios 36, 42, 43, 45, 49, 51, and 52 were insightful. Scenario 36 proved that taxi time allocated to Passenger Two was necessary to maximum the overall gathered fare values despite all the circumstances being larger in the Passenger One parameters. For Scenarios 42 and 43 the optimum solution was shown to be to assign all of the taxi time to pursue Passenger One. Scenarios 42 and 43 indicate that the conditional fare value is also important but that the hail rate is just as important to optimize the use of the taxi. For Scenarios 45 and 49 the complex relationship between hail rates and conditional fare values were apparent as the time allocation was balanced between the passengers. The examination of Scenarios 51 and 52 reveals all the taxi time should be spent on Passenger One. These circumstances are an example of when the conditional fare value over compensates for the small probabilities of successfully picking up the passenger.

The numerical analysis and Monte Carlo simulation results can be applied to important national assets. Appendix A presents such a case. The appendix is classified TOP SECRET. To obtain a copy of this classified appendix, please contact the Naval Postgraduate School's Dudley Knox Library.

B. FURTHER RESEARCH

Further research areas include considering the conditional fare values, I_1 and I_2 , as random variables; add another task to be accomplished during the assumed available time and choose PDFs for the rates that match more closely observed variations. A last option to further research is to apply this thesis to actual practice.

An area to research would be to consider the conditional fare values, I_1 and I_2 , as random variables along with the distributed hail rates, R_1 and R_2 . This must be done to ensure that these methods are of value when the conditional fare values are not known precisely, which is likely the realistic case. Fortunately, the concepts used in this thesis, to show that these methods are of value when the hail rates are not known precisely, can be applied directly to this problem.

Another avenue of study could be to include three or more tasks with independent uniformly distributed random variables for the rates and constants for the conditional fare values. This would involve slight modifications to the equations of numerical analysis and Monte Carlo simulation to find the expected optimum time allocation for each task and could extend the findings.

Another area for further research is to improve upon the accuracy of the PDFs for the independent uniformly distributed random variable rates used in numerical analysis. Using historical data should yield the appropriately estimated PDFs.

The final recommendation for further research is to put this model into actual practice and run an analysis on the amount of total fare gathered. The transition to practice may indicate that improvements are needed to the theory or mathematical considerations that were not initially considered and thus furthering the applicability of this model.

APPENDIX A. CLASSIFIED CHAPTER

Classified TOP SECRET. Contact Naval Postgraduate School's Dudley Knox Library.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. EQUATION (1.1) AND (1.2) DERIVATION

This appendix illustrates how Equations (1.1) and (1.2) are derived [9].

Let I represent the total expected fare for a taxi that attempts to collect two fares, from Passenger One and Passenger Two. I_1 and I_2 are the fares that will be collected from Passengers One and Two assuming the taxi picks them up. These fares, are conditioned on the fact that the passengers are picked up, are called the conditional fares. The expected fare collected from passenger one is

$$\begin{aligned} & E(\text{fare from Passenger One}) \\ &= I_1 \Pr(\text{Passenger One is picked up}) + 0 \Pr(\text{Passenger One is not picked up}) \quad (\text{B.1}) \\ &= I_1 \Pr(\text{Passenger One is picked up}). \end{aligned}$$

The number of passengers, N , serviced in a fixed amount of time can be considered a count of random events, and therefore is modeled as a Poisson random variable. A Poisson random variable has probability mass function

$$\Pr(N = n) = \frac{(RT)^n \exp(-RT)}{n!} \quad \text{for } n = 0, 1, 2, \dots \quad (\text{B.2})$$

where R is the average rate of passengers picked up and time T is the time spent looking for passengers. Therefore, the probability of no passengers being picked up in time T is $\Pr(N = 0) = \exp(-RT)$. Therefore, the probability that one or more passengers are picked up is

$$\Pr(N > 0) = 1 - \Pr(N = 0) = 1 - \exp(-R_1 T_1). \quad (\text{B.3})$$

Passenger One is the first passenger, so he/she is picked up if $N > 0$, i.e.,

$$\Pr(\text{Passenger One is picked up}) = \Pr(N > 0) = 1 - \exp(-R_1 T_1) \quad (\text{B.4})$$

where R_1 is the average rate of Passenger One being picked up and time T_1 is the time spent looking for Passenger One. In our case, the fare is earned for the first passenger and no fare is earned for the others, so, using Equations (B.1) and (B.4), the expected fare from Passenger One is

$$E(\text{fare from Passenger One}) = I_1 [1 - \exp(-R_1 T_1)]. \quad (\text{B.5})$$

Similarly, the expected fare from Passenger Two is

$$E(\text{fare from Passenger Two}) = I_2[1 - \exp^{(-R_2 T_2)}] \quad (\text{B.6})$$

where R_2 is the average rate of Passenger Two being picked up and time T_2 is the time spent pursuing Passenger Two. If the total time spent looking for both Passengers One and Two is $T_{total} = T_1 + T_2$, then Equation (B.6) can be written in terms of T_1 , i.e.,

$$E(\text{fare from Passenger Two}) = I_2 \left\{ 1 - \exp^{[-R_2(T_{total} - T_1)]} \right\}. \quad (\text{B.7})$$

Therefore, using Equations (B.5) and (B.7), the total expected fare during time T_{total} is

$$E(\text{total fare collected}) = I_1[1 - \exp^{(-R_1 T_1)}] + I_2[1 - \exp^{[-R_2(T_{total} - T_1)]}]. \quad (\text{B.8})$$

In order to maximize his/her expected total fare collected, the smart taxi driver will choose the time to spend looking for Passenger One, T_1 , and the time to spend looking for Passenger Two, $T_2 = T_{total} - T_1$, such that the derivative of Equation (B.8) with respect to T_1 at that value of T_1 is zero, provided there exists a T_1 between zero and T_{total} that makes the derivative of Equation (B.8) equal to zero. Let the value of T_1 that maximizes Equation (B.8) be denoted by T_1^* . The taxi driver will maximize his expected total fare collected subject to the constraint $0 \leq T_1 \leq T_{total}$ if the time he spends looking for Passenger One is

$$T_1 = \begin{cases} 0 & \text{if } T_1^* < 0 \\ T_1^* & \text{if } 0 \leq T_1^* \leq T_{total} \\ T_{total} & \text{if } T_1^* > T_{total} \end{cases}. \quad (\text{B.9})$$

In the remainder of this appendix, we solve for T_1^* , the value of T_1 that maximizes Equation (B.8). Differentiating Equation (B.8) with respect to T_1 yields

$$0 = I_1 R_1 e^{(-R_1 T_1^*)} - I_2 R_2 e^{[-R_2(T_{total} - T_1^*)]}. \quad (\text{B.10})$$

Solving this equation for T_1^* yields

$$T_1^* = \frac{\ln\left(\frac{I_1 R_1}{I_2 R_2}\right) + R_2 T_{total}}{(R_2 + R_1)} . \quad (B.11)$$

Therefore the taxi cab driver will maximize his expected collected fare if (s)he allocates his time such that the time he spends looking for passenger one is T_1 as calculated using Equations (B.9) and (B.11).

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. MATLAB CODE FOR NUMERICAL ANALYSIS

MATLAB Method I, numerical analysis

By Crystal Warren with assistance from Cole Johnson and Frank E. Kragh.

%% Beginning of Method I code.

```
function y = uniformv4(R1, R2, R1min, R1max, R2min, R2max)
% Everything on the right side is what you are inputting into this
function,
% everything on the left(i.e. 'y') is what the equation will output.
```

```
H = 1/ ( ...
        (R1max-R1min) ...
        * (R2max-R2min) ...
    );
% Height (H) represents the height of the volume that two variables (R1
and
% R2) with a function make of a two variable PDF. R1 parameters are
"para"
% of (1,1) and (1,2). R2 parameters are "para" of (2,1) and (2,2). This
% equation originates from finding the volume of two variables of a
% PDF = 1 = Height*(point12-point11)*(point22-point21)
inside = (R1 < R1max) & (R1 > R1min) & (R2 < R2max) & (R2 > R2min);
% the above equation creates a matrix that shows only where R1 and
R2
% is true
y = H*inside;
% This results in a matrix where each entry is 0 or H.
```

```
function [ y ] = theintegrandv5( R1, R2 )
%% Time (time constant)
timespan = 1; %hours

[R1min, R1max, R2min, R2max, CIV1, CIV2] = getRateMaxMinv2();
(function [R1min, R1max, R2min, R2max, CIV1, CIV2] = getRateMaxMinv2()
```

```
R1min = 1;
R1max = 10;
```

```
R2min = 1;
R2max = 10;
```

```
CIV1 = 100; % CIV = Conditional Intelligence Value
CIV2 = 100;)
```

```
uniformParameters = [R1min R1max; R2min R2max];
```

```
time_optimum1 = ( log( ...
                    (CIV1*R1)./(CIV2*R2) ...
                ) ...
                + R2*timespan ...
```

```

) ./ ...
(R1+R2) ; % This is Equation (1.1).

time_optIsZero = time_optimum1 <= 0;
time_optIsTimespan = time_optimum1 >= timespan;
time_optIsZeroOrTimespan = time_optIsZero | time_optIsTimespan;

% Expected Intelligence Values = EIV = CIV1*(1-exp(-R1*T1)) + CIV2*(1-
exp(-R2*T2))
EIV0 = CIV2*(1-exp(-R2*timespan)); % EIV at T1 = 0;
EIVatTimespan = CIV1*(1-exp(-R1*timespan)); % EIV at T1 = timespan
EIVatTimespanIsBigger = EIV0 < EIVatTimespan; % which endpoint has
larger
%EIV

time_optimum1(time_optIsZeroOrTimespan) =
EIVatTimespanIsBigger(time_optIsZeroOrTimespan)*timespan;

f_R1R2_r1r2 = uniformv4(R1, R2, R1min, R1max, R2min, R2max);

y = time_optimum1.*f_R1R2_r1r2;

[R1min, R1max, R2min, R2max, CIV1, CIV2] = getRateMaxMinV2();
timespan = 1;

expectation = quad2d(@theintegrandv5, R1min, R1max, R2min, R2max)
% expectation represents the average T1 value for Method I.

```

APPENDIX D. MATLAB CODE FOR MONTE CARLO SIMULATION

MATLAB Method II, Monte Carlo Simulation

By Crystal Warrenne with assistance from Cole Johnson and Frank E. Kragh.

```
%% Method II is to check the results of Method I
(TwotargetsMethodlv4FEK.m
%% + theintegrands5.m + uniformv4.m) value for the expectation should
match
%% the value of T1 in this code.
```

```
% clear
close all
clc
clear all
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%
```

```
[R1min, R1max, R2min, R2max, CIV1, CIV2] = getRateMaxMinV2();
%% Allows for the variable inputs to be inputted once and ran in both
%% Methods I and II.
(function [R1min, R1max, R2min, R2max, CIV1, CIV2] = getRateMaxMinV2()
```

```
R1min = 1;
R1max = 10;

R2min = 1;
R2max = 10;
```

```
CIV1 = 100; % CIV = Conditional Intelligence Value
CIV2 = 100;)
```

```
%% Number of rolls; K must be a range that is so large that it does not
%% affect the outcome regardless of value.
k = 1e6 ;
```

```
R1 = R1min + (R1max-R1min)*rand(k,1);
R2 = R2min + (R2max-R2min)*rand(k,1);

timespan = 1;

time_taken = ...
    ( (log(CIV1)+log(R1) - log(CIV2)-log(R2)) ...
      + R2*timespan ...
    ) ...
    ./ (R1+R2); % Equation (1.2) is written differently to ensure
the
    % same results are calculated.
```

```
time_takenNotRight = (time_taken <= 0) | (time_taken >= timespan);
```

```

    % This limits the results of T1 to be between 0 and the given value
of
    % timespan.

    TIVatZip = CIV2*(1-exp(-R2*timespan));
    TIVatTimespan = CIV1*(1-exp( - R1*timespan ) );

    %%%%%%%%%%%%%%%
    % Need a section that compares the above two TIVs and chooses
optimum
    % time (0 or Timespan) which is the desired answer when time_taken
is
    % not right.

    %for each k, find which endpoint has larger TIV
    TIVatTimespanIsLarger = TIVatZip < TIVatTimespan;

    %for each k that was not right, let time_taken be that endpoint
    time_taken(time_takenNotRight) =
TIVatTimespanIsLarger(time_takenNotRight)*timespan;

    %%%%%%%%%%%%%%%
    % So run a test for a timetaken-is-not-right-case to look at plot
of
    % TIV verses time allocated for task1
    timespan_graph = [0.0:0.01:timespan]';

    R1bad = R1(time_takenNotRight);

    R2bad = R2(time_takenNotRight);
    if ( length(R1bad) > 0 )

        indexx = randi(length(R1bad));
        x = time_taken(time_takenNotRight);
        y = x(indexx);

        TIV = CIV1*(1-exp( - R1bad(indexx)*timespan_graph ) ) ...
            + CIV2*(1-exp(-R2bad(indexx)*(timespan-timespan_graph)));

        TIVAvg = mean(TIV);
        plot(timespan_graph,TIV)
        text(.3,TIVAvg,['optimum time assignment is ' num2str(y)])
    else
        disp(['All calculated times for task1 are in (0, '
num2str(timespan) ')'])
    end

    time_2 = timespan - time_taken;

    % figure('Name','Histogram of time_taken')
    % hist(time_taken,1e3)
    % title('Combined Histogram')

```



```
% figure('Name','Histogram of time_2')
% hist(time_2,1e3)
% title('Combined Histogram')

%% Statistics

disp(['Stats for T1'])
average = mean(time_taken)

disp(['Stats for T2'])
average = mean(time_2)
```

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. PROBABILITY THAT HAIL RATE OF PASSENGER ONE IS LARGER THAN THE HAIL RATE OF PASSENGER TWO

In this appendix, the probability used in IV.A.1 is calculated. Specifically, we calculate $\Pr(R_1 > R_2)$ given that R_1 and R_2 are independent uniformly distributed random variables with joint probability density function

$$f_{R_1 R_2}(r_1, r_2) = \begin{cases} \left(\frac{1}{216}\right)hr^2 & \text{if } 1/hr \leq r_1 \leq 25/hr \text{ and } 1/hr \leq r_2 \leq 10/hr \\ 0 & \text{otherwise.} \end{cases} \quad (\text{E.1})$$

The probability can be calculated as

$$\begin{aligned} \Pr(R_1 > R_2) &= \iint_{r_1 > r_2} f_{R_1 R_2}(r_1, r_2) dr_1 dr_2 \\ &= \left(\frac{1}{216}\right)hr^2 (Area_{trapezoid}) \end{aligned} \quad (\text{E.2})$$

where the $Area_{trapezoid}$ is the area of the trapezoid with height (10 per hour minus 1 per hour) = 9 per hour, one base of the trapezoid is (25 per hour minus 1 per hour) = 24 per hour, and the other base of the trapezoid is (25 per hour minus 10 per hour) = 15 per hour.

$$Area_{trapezoid} = \left(\frac{1}{2}\right)\left(\frac{9}{hr}\right)\left(\frac{24}{hr} + \frac{15}{hr}\right) = \frac{175.5}{hr^2}. \quad (\text{E.3})$$

Therefore,

$$\Pr(R_1 > R_2) = \left(\frac{1}{216}\right)hr^2 \left(\frac{175.5}{hr^2}\right) = .8125. \quad (\text{E.4})$$

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

- [1] Justin Luther, “UPS Accelerates Use of Routing Optimization Software to Reduce 100 Million Miles Driven” in *UPS Pressroom*. [Online]. Available: <https://www.pressroom.ups.com/pressroom/ContentDetailsViewer.page?ConceptType=PressReleases&id=1426329559785-791>. Accessed August, 3, 2016.
- [2] “Map,” n.d. Google [Online]. Available at: <http://images.google.com/>. Accessed: Jan. 29, 2016.
- [3] “Taxi,” *Dreamstime*. [Online]. Available at: <http://www.dreamstime.com/royalty-free-stock-images-cartoon-yellow-taxi-car-image21652969>
- [4] W. Mendenhall. *Introduction to Probability and Statistics*. 7th ed. Boston: PWS, 1987, pp. 139,140, 149–152.
- [5] T.W. Anderson and Jeremy D. Finn. *The New Statistical Analysis of Data*. 1st ed. New York: Springer-Verlag, 1996, pp. 23, 247, 326.
- [6] Richard J. Larsen and Morris L. Marx. *An Introduction to Mathematical Statistics and Its Applications*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2001, pp. 30–31, 247, 140–150, 161,51, 190–193,167-169.
- [7] B. Dutertre. “Dynamic scan scheduling” in *Institute of Electrical and Electronics Engineers Real-Time Systems Symposium*, Austin, TX ‘02 23rd IEEE symposium. December 2-3, 2002, pp. 327–336.
- [8] C. Dukes. *Optimizing the Signals Intelligence Tasking Process*, M.S. thesis, Graduate School of Engineering and Applied Sciences, Naval Postgraduate School, Monterey, CA, 2010.
- [9] K. St. Germain. *Optimizing the Signals Intelligence Resource Allocation Process*, M.S. thesis, Graduate School of Engineering and Applied Sciences, Naval Postgraduate School, Monterey, CA, 2013.
- [10] R. C. Larson and A. Odoni, n.d. 2.10 Probability Density Functions, [Online]. Available at: http://web.mit.edu/urban_or_book/www/book/chapter2/2.10.html. Accessed January 29,-2016.
- [11] L. Gonick and W. Smith, *The Cartoon Guide to Statistics*, 1st ed. New York: Harper, 2005, pp. 53.

- [12] E. Waltz. “Explicit Models of the Targets of Analysis” in *Quantitative Intelligence Analysis*. 1st ed. Lanham, MD: Rowman & Littlefield, 2014, ch. 7, sec. Methodology and Technology Challenges, pp. 185, 186, 193.
- [13] K. Binder and D. Heermann, *Monte Carlo Simulation in Statistical Physics*. Berlin: Springer, 2002, pp. 49–51.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California